Linear Algebra

Samira Hossein Ghorban
s.hosseinghorban@ipm.ir

Fall, 2021

Review: The Gram-Schmidt process

- The Gram-Schmidt process
  1. starts with independent vectors $v_1, \ldots, v_n$
  2. ends with orthonormal vectors $q_1, \ldots, q_n$.

- At step 1: $q_1 = \frac{1}{\| v_1 \|} v_1$.

- At step j ($2 \leqslant j \leqslant n$):

  1. it subtracts from a_j its components in the directions $q_1, \ldots, q_{j-1}$ that are already settled:

$$Q_j = v_j - \langle v_j, q_1 \rangle\, q_1 - \cdots - \langle v_j, q_{j-1} \rangle\, q_{j-1}.$$

  2. $q_j = \frac{1}{\| Q_j \|} Q_j$.

  3. $\mathrm{span}(\{v_1, \ldots, v_j\}) = \mathrm{span}(\{q_1, \ldots, q_j\})$.

The factorization $A = QR$

- The Gram-Schmidt process
  1. starts with independent vectors $A_1, \ldots, A_n$, consider these vectors as columns of a matrix A.
  2. ends with orthonormal vectors $q_1, \ldots, q_n$, consider these vectors as columns of a matrix Q.
- What is the relation between these matrices A and Q?
- Think about three vectors A_1, A_2, A_3.
- $\text{span}(\{A_1, A_2, A_3\}) = \text{span}(\{q_1, q_2, q_3\})$.
- The idea is to write the A's as combinations of the q's.

$$
\begin{aligned}
A_1 &= \langle A_1, q_1 \rangle q_1 \\
A_2 &= \langle A_2, q_1 \rangle q_1 + \langle A_2, q_2 \rangle q_2 \\
A_3 &= \langle A_3, q_1 \rangle q_1 + \langle A_3, q_2 \rangle q_2 + \langle A_3, q_3 \rangle q_3.
\end{aligned}
$$

The factorization $A = QR$

- By

$$\begin{aligned}
A_1 &= \langle A_1, q_1\rangle\, q_1 && +0q_2 && +0q_3\\
A_2 &= \langle A_2, q_1\rangle\, q_1 && +\langle A_2, q_2\rangle\, q_2 && +0q_3\\
A_3 &= \langle A_3, q_1\rangle\, q_1 && +\langle A_3, q_2\rangle\, q_2 && +\langle A_3, q_3\rangle\, q_3,
\end{aligned}$$

- We obtain:

$$\underbrace{\begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \langle A_1, q_1\rangle & \langle A_2, q_1\rangle & \langle A_3, q_1\rangle \\ 0 & \langle A_2, q_2\rangle & \langle A_3, q_2\rangle \\ 0 & 0 & \langle A_3, q_3\rangle \end{bmatrix}}_{R}$$

QR factorization

- In QR factorization for the matrix A (with independent columns):
  1. the first factor Q has orthonormal columns.
  2. R is upper triangular (The second factor is called R, because the nonzeros are to the right of the diagonal).
- Example:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix} = QR$$

Simple example

- This failure is almost certain when there are several equations and only one unknown:

$$\begin{cases} 2x = b_1 \\ 3x = b_2 \\ 4x = b_3 \end{cases}$$

- In spite of their unsolvability, inconsistent equations arise all the time in practice. They have to be solved!
- One possibility is to determine x from part of the system, and ignore the rest; this is hard to justify if all m equations come from the same source.
- Rather than expecting no error in some equations and large errors in the others, it is much better to choose the x that minimizes an average error E in all m equations.

Squared error

- It is much better to choose the x that minimizes an average error E in the m equations:

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2$$

- If there is an exact solution, the minimum error is $E = 0$
- In the more likely case that b is not proportional to a We ***solve*** $Ax = b$ by minimizing

$$E^2 = \|ax - b\|^2.$$

- The best estimation:

$$\hat{x} = \frac{a^T b}{a^T a}.$$

Least squares problems with several variables

- So, instead of one column and one unknown x, the matrix now has n columns. The number m of observations is still larger than the number n of unknowns.

- So it must be expected that $Ax = b$ will be inconsistent. Probably, there will not exist a choice of x that perfectly fits the data b.

- In other words, the vector b probably will not be a combination of the columns of A; it will be outside the column space.

- Solution: finding $\hat{x}$ which minimizes $E = \|b - Ax\|$.

- Equivalently, finding $\hat{x}$ such that the error vector $e = b - A\hat{x}$ be perpendicular to that space $C(A)$.

Least squares problems with several variables

- Finding $\hat{x}$ such that the error vector $e = b - A\hat{x}$ be perpendicular to that space $C(A)$ is equivalent to

$$A^T(b - A\hat{x}) = 0$$
$$A^TA\hat{x} = A^Tb$$

- If A^TA is invertible, then the best square estimation :

$$\hat{x} = (A^TA)^{-1}A^Tb.$$

Matrix A^TA

- The matrix A^TA is certainly symmetric.
- $N(A^TA) = N(A)$.
- If A has independent columns, then A^TA is invertible.
- We have shown that the closest point of the column space of A to b is
$$p = A(A^TA)^{-1}A^Tb.$$
- The matrix that gives p is a projection matrix, denoted by P:
$$P = A(A^TA)^{-1}A^T.$$

Projection matrix

- This matrix has two main properties:

  1. P is a symmetric matrix.
  2. Its square is itself, again: $P^2 = P$.

Projection matrix

- This matrix has two main properties:

    1. P is a symmetric matrix.

    2. Its square is itself, again: $P^2 = P$.

- Conversely, any symmetric matrix with $P^2 = P$ represents a projection.

Least-Squares Fitting of Data

- Suppose that we do a series of experiments, and expect the output b to be a linear function of the input t. We look for a straight line $b = C + Dt$.
- To measure the distance to a satellite on its way to Mars!
- We vary the load on a structure, and measure the movement it produces.

Example

- Assume three measurements $b_1 = 1, b_2 = 1, b_3 = 3$ for $t = -1, t = 1, t = 2$, respectively.
- Thus, every $C + Dt$ would agree exactly with b:

$$\begin{cases} C - D = 1 \\ C + D = 1 \\ C + 2D = 3 \end{cases}$$

- What is the best estimation?

Weighted Least Squares

- Now suppose that the two observations are not trusted to the same degree.

- A simple least-squares problem is to estimate two observations $x = b_1$ and $x = b_2$.
- if $b_1 \neq b_2$, then we are faced with an inconsistent system of two equations in one unknown:

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

- Thus

$$E^2 = (x - b_1)^2 + (x - b_2)^2$$

- The best square estimation:

$$\hat{x} = \frac{b_1 + b_2}{2}.$$

Weighted Least Squares

- Now, suppose that the the value $x = b_1$ may be obtained from a more accurate source from a larger $x = b_2$.
- The simplest compromise is to attach different weights w_1 and w_2 and choose the $\hat{x}_W$ that minimizes the weighted sum of squares:

$$E^2 = w_1(x - b_1)^2 + w_2(x - b_2)^2.$$

- The best square estimation :

$$\hat{x} = \frac{w_1^2 b_1 + w_2^2 b_2}{w_1^2 + w_2^2}.$$

- Generally:

$$(A^T W^T W A)\hat{x}_W = A^T W^T W b.$$

Thank You!